EMOTIONAL IDENTITY OF MOVIES

Luca Canini, Sergio Benini, Pierangelo Migliorati and Riccardo Leonardi

DEA-SCL, University of Brescia, Via Branze 38, 25123, Brescia, Italy
Email: {*firstname.lastname*}@*ing.unibs.it*

ABSTRACT

In the field of multimedia analysis, attempts that lead to an emotional characterization of content have been proposed. In this work we aim at defining the emotional identity of a feature movie by positioning it into an emotional space, as if it was a piece of art. The multimedia content is mapped into a trajectory whose coordinates are connected to filming and cinematographic techniques used by directors to convey emotions. The trajectory evolution over time provides a strong characterization of the movie, by locating different movies into different regions of the emotional space. The ability of this tool in characterizing content has been tested by retrieving emotionally similar movies from a large database, using IMDb genre classification for the evaluation of results.

***Index Terms*—** Emotional identity, Video retrieval

1. INTRODUCTION

In the movie industry, beside big film studios, smaller independent companies are growing since they are able to perform all the necessary movie editing by using only a digital camcorder and a computer. A similar situation can be observed in the amateur domain: everyone can produce his/her own home video, edit it with a personal computer and share it thanks to dedicated web sites. Consequently, since the amount of multimedia content is increasing, there is a need to organize it so as to efficiently retrieve relevant material from a database.

In order to answer these requests, one of the strategies is to extract the information present in a multimedia content at the semantic level. Two different approaches have been adopted: cognitive and emotional [1]. The cognitive approach analyses the audiovisual sequence, trying to understand what is happening in a scene in terms of location, characters and events. These methods are very spread and there is a huge amount of related work (see for example [2]). On the other side there is the emotional approach, first discussed in detail in [1], whose aim is to characterize a video trying to define emotions it communicates. This approach has been less investigated, but its importance is rapidly growing.

Speaking about emotions, we must keep in mind that they are personal: everyone reacts to an event in a way that depends on education, personal experiences and many other factors. Therefore, it seems impossible to define an objective method to find emotions in videos. However, talking about movies, a concrete way to do this is using the *expected mood* [3], *i.e.*, emotions the film-maker intended to communicate when he produced the movie for a particular audience with a common cultural background. So far, the emotional characterization has been used to study a narrow set of situations, like sport events [4] or movies that belong to a particular genre [5]. The method presented in [6] tries to map a video into a limited set of emotions by using *HMM*. However, the emotional model of *pleasure-arousal* [1], which is common to most of these approaches, is often too restrictive for characterizing the emotional complexity of a video sequence.

In this paper instead, we consider a movie as a piece of art and we place it into a space which is similar to those used for defining the emotional identity of design objects [7]. In order to build our *emotional space* we directly investigate the cinematographic techniques adopted by directors to convey emotions. As a result, the movie is mapped into a trajectory, using some relevant features connected to audiovisual filming techniques. The evolution of the trajectory over time provides a strong characterization, since emotionally different movie sequences occupy different regions of the emotional space. In the experimental part, this analysis tool is used for retrieving movies with similar emotional identity from a large database.

This document is organized as follows: in Section 2 the emotional space is presented, while the characterization of its three axes is described in Section 3. Section 4 discusses how a movie can be represented in the proposed space. In Section 5 our framework is tested for retrieving emotionally similar movies. Conclusions are finally drawn in Section 6.

2. SPACE OF EMOTIONAL IDENTITY

One important task during the creation of any design object is to define its emotional identity. In [7] the author presents a tool to describe it, by placing the product in a 3D emotional space according to its shape, efficiency and social context. In this space the three axes refer to the so called *natural, temporal* and *energetic* dimension, respectively. The natural one splits the space into a passional hemisphere, referred to warm emotions, and a reflective hemisphere, that represents offish and cold feelings. The temporal axis divides the space into two zones, one related to the slowness of the past and experience and the other describing a character quickly projected towards the future. Finally, the energetic axis identifies objects with high emotivity and, in opposition, minimal ones.

Considering a movie as a piece of art and design, we transpose the emotional identity in cinematographic terms. At first we associate each axis to a couple of adjectives in a dichotomic relationship. To the natural axis we link the couple *warm/cold*. The temporal one is described in terms of *dynamic/slow*, while the dichotomy *energetic/minimal* is associated to the third axis. Then, we look for filming and editing techniques used by directors to convey emotions which we associate to the selected dichotomies (see Section 3).

In the defined space, a movie is represented by a trajectory that describes the temporal evolution of its emotional identity (as shown in Figure 1). This trajectory, free to move over the entire space, gives an accurate characterization of the movie, since it is not restricted to a fixed set of emotions or to a pleasure-arousal scheme.

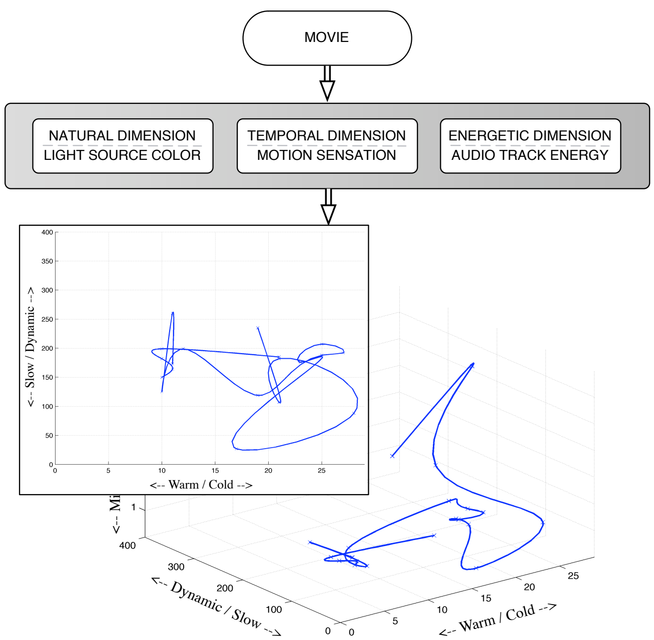

Fig. 1: Space of emotional identity: general framework (top) and trajectory from an excerpt of "Matrix" (bottom).

3. CONVEY EMOTIONS THROUGH THE SCREEN

In order to convey emotions and give a scene a specific mood, directors have at their disposal lots of techniques [8], such as the use of colour (of objects and for the scene illumination), the shot type (*e.g.*, long, medium, close-up), the aesthetic organization of the scene, the accompanying audio, etc.

To define the axes of our space, we link the associated dichotomies with specific filming and editing techniques. For the natural axis we consider the colour of the light source illuminating the scene, for the temporal one we select the motion sensation evoked in the movie, and finally we associate the energy of the audio track to the energetic axis.

3.1. Light source colour

The spectral composition of the light source is very important in the definition of a scene character. Objects have not their own colours, which are instead due to the interaction with the incident electromagnetic radiation. Consequently, colours seen by the audience watching a movie are influenced by the light source, also called illuminant, used by directors during scene recording. For the natural axis we consider the dichotomy warm/cold and we associate it to the illuminant value. Humans can distinguish between warm colours (red, yellow, etc.) and cold ones (green, blue, etc.) [9]. For example, in Figure 2 the frame on the left shows a scene with a yellow polarised illuminant which evokes a pleasant sensation, while the one on the right suggests a quite offish feeling because of the grayish illuminant.

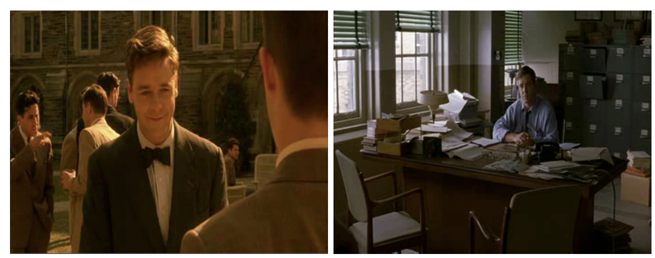

Fig. 2: Two frames from "A beautiful mind". The left frame evokes a warm sensation, the other an offish feeling.

Given a movie, we estimate the illuminant colour for each shot by improving a *white patch* algorithm [10]. Let $I(\lambda)$ be the intensity of the light source, λ its wavelength and $s(\lambda)$ the reflectance of a single point in a scene whose colour f_i $(i = 1, 2, 3)$ is:

$$\begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix} = \int_\lambda I(\lambda)s(\lambda) \begin{bmatrix} c_1(\lambda) \\ c_2(\lambda) \\ c_3(\lambda) \end{bmatrix} d\lambda \qquad (1)$$

where the c_i are the *colour matching functions* of the given colour space. If a point is a "truly white point", *i.e.*, it has maximum reflectance for each λ, its colour is an accurate estimation of the scene illuminant. This is true under the mostly verified hypothesis that the light source is spread quite uniformly over the scene. Anyway, in order to make a robust estimation, we consider a minimum "truly white area" and we average results on single frames for each shot.

A novel procedure to map the components f_i of the illuminant on a one-dimensional warm/cold scale (natural axis) is then proposed. The Black Body radiation, whose spectral composition depends only on temperature (Figure 3, top), provides a suitable starting point. However, this radiation has not green hues and the chromatic distance between its points is not linear with temperature. To solve these problems, we first build a dummy radiation by switching the position of the green and the blue channels. Then, we linearise both the original and the dummy radiation. Finally, combining these results with an appropriate non-uniform quantization law, we map the illuminant on the $\mathcal{N}$-th interval of the axis (Figure 3, bottom).

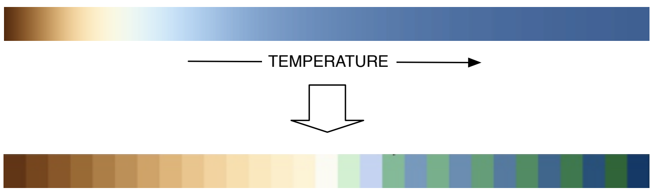

Fig. 3: From the Black Body radiation (top) to the built natural axis (bottom).

3.2. Motion dynamics

Temporal dynamics are often employed by directors to stress the emotional identity of a scene. To transmit the audience a sensation of speed and dynamism or a feeling of calm and tranquility, directors can rely on the shot pace and type, the camera motion and the movement of objects. Thus, for the temporal axis we consider the dichotomy dynamic/slow and we bind it to the motion dynamics.

A shot is mapped on this axis using an index $\mathcal{T}$ given by:

$$\mathcal{T} = \left[2 - min\left(2, \frac{l_{shot}}{l_{avg}}\right)\right] + k \cdot \mathcal{M} \qquad (2)$$

The first term of $\mathcal{T}$ is related the shot length l_{shot} and uses the average shot length ($l_{avg} = 5.02s$) computed on a large database of movies produced on the last twenty years [11]. A short shot presents a big value of $\mathcal{T}$, since short shots convey high narration pace. Moreover, if l_{shot} is at least double than l_{avg}, this contribution becomes null. The second term is connected to the motion activity. It captures the intuitive notion of intensity of action, not distinguishing between camera and object motion, and it is measured by the standard deviation of motion vector modules. This term is then averaged over the entire shot, obtaining $\mathcal{M}$, and normalized to the same scale as the previous term by a coefficient k.

3.3. Audio track log-energy

Important scenes in movies are usually stressed thanks to a particular choice of the soundtrack, *e.g.,* gentle and pleasant music for a romantic moment, loud and aggressive audio for an action sequence, silences and reprises in a dialogue. For the energetic axis we use the dichotomy energetic/minimal and we associate it to the energy of the audio track.

Log-energy $\mathcal{E}$ is computed for each shot by using a 8 kHz single-channel audio signal. To highlight the presence of brief and intense events (like thunders, screams, etc.), only audio samples above an adaptive threshold are taken into account.

4. DRAWING A MOVIE IN THE EMOTIONAL SPACE

In the defined emotional space, a movie is drawn as a cloud of points, where each point, defined by a triplet $\{\mathcal{N}, \mathcal{T}, \mathcal{E}\}$, represents a shot. During the movie playback, these points are connected in temporal order by a cubic spline, creating a trajectory which describes the evolution of the movie emotional identity.

4.1. Solid summaries

Drawing the trajectory of an entire movie may result in a too complex description of its emotional identity. A higher level view is provided by a 3D-solid that summarises the fundamental characteristics of the trajectory (Figure 4).

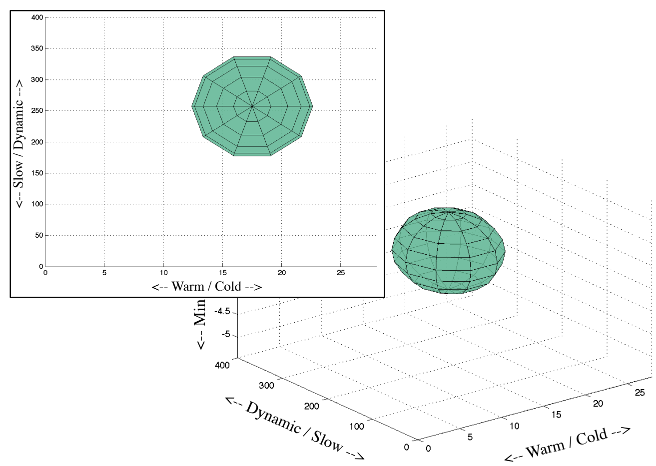

Fig. 4: A 3D-solid built on the whole trajectory provides a synoptic overview of the emotional identity of "Matrix".

The solid *colour* is the average illuminant value, while its *geometric shape* is defined by the smoothness of the trajectory: the smoother the trajectory, the smoother the solid surface. The solid *centroid* c is obtained by averaging the shot positions, while solid *dimensions* σ_j are computed as the standard deviations of the shot triplets $\{\mathcal{N}, \mathcal{T}, \mathcal{E}\}$ over the three axes. Based on these parameters, it is possible to define an "emotional distance" D between the solids $\mathcal{S}_A$ and $\mathcal{S}_B$ representing two movies:

$$D(\mathcal{S}_A, \mathcal{S}_B) = \sum_{j=1}^{3} \alpha_j |c_{A,j} - c_{B,j}| + \frac{1}{\beta} \sum_{j=1}^{3} \alpha_j |\sigma_{A,j} - \sigma_{B,j}|$$

where coefficients α_j are used to normalize the axes to a common scale, while β adequately weighs the two terms.

5. RETRIEVAL BY EMOTIONAL IDENTITY

Our framework has been tested by retrieving emotionally similar content from a database of 87 movies. As a ground-truth we considered for each movie the first two cinematographic genres given by IMDb [12]. To build the similarity matrix shown in Figure 5, movies have been ordered by their genres, by the incidental presence of sagas, and by their emotional affinity judged by a human observer. This reveals that movies sharing the same genre are usually at small emotional distance (*i.e.,* blue color), such as the "Action" movies which cluster at the top-left of the matrix.

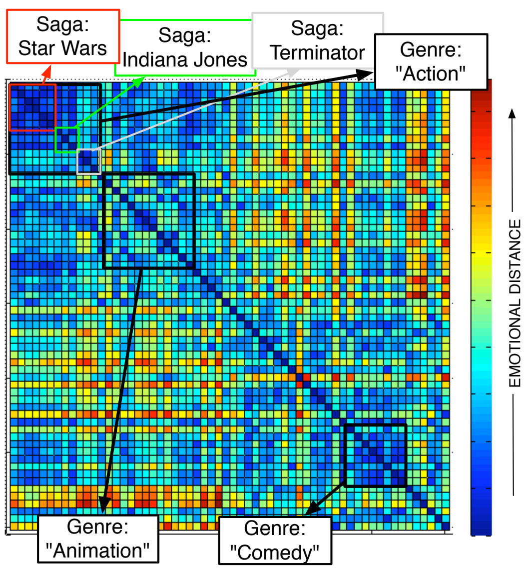

Fig. 5: Similarity matrix based on emotional distances.

Starting from these considerations, we verified the efficiency of our approach by building an application of movie retrieval based only on their emotional identity. Given a query movie, the application is able to retrieve from the database those movies whose 3D-solids are at lower emotional distance from the query one. The system performance is evaluated in terms of Precision-Recall measured with respect to the IMDb genres of the retrieved movies. Curves in Figure 6 are averaged on the results obtained considering all single movies as queries. As shown in this figure, the combination of the three emotional axes is better performing, in terms of P-R curves, than the systems employing single axes only.

To the best of our knowledge, the proposed system is the first that uses an emotional characterization for retrieving videos. To assess its full potential, further work for integration in a modern content-based retrieval system is needed.

6. CONCLUSIONS

In this work we treat a movie as a piece of art, by defining its emotional identity and positioning it into an emotional space.

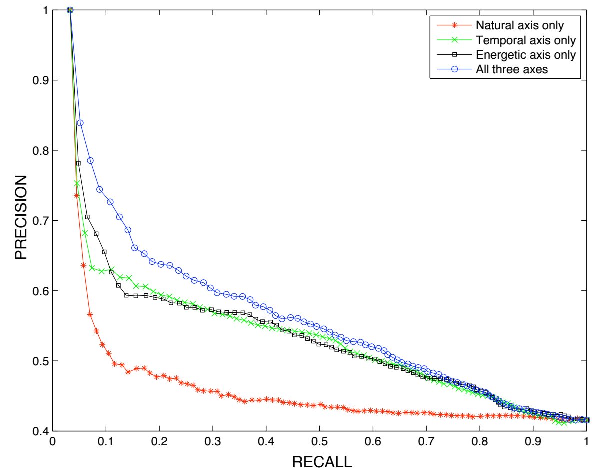

Fig. 6: Precision-Recall curves of the retrieval system.

The emotional characterization of the movie over time is expressed as a trajectory whose behavior is determined by filming techniques based on the use of colour, motion and audio. The obtained characterization has been tested for retrieving emotionally similar movies from a large database. Obtained results, though preliminary, are encouraging and suggest that the emotional approach to video retrieval, rather than totally alternative, should be considered at least as complementary to traditional retrieval and recommender systems.

7. REFERENCES

[1] A. Hanjalic and L.-Q. Xu, "Affective video content representation and modeling," *IEEE Transactions on Multimedia*, vol. 7, no. 1, pp. 143–154, Feb. 2005.

[2] S.-F. Chang, W.-Y. Ma, and A. Smeulders, "Recent advances and challenges of semantic image/video search," in *Proc. of ICASSP*, Hawaii, USA, Apr. 2007.

[3] A. Hanjalic, "Extracting moods from pictures and sounds," *IEEE Signal Processing Magazine*, vol. 23, no. 2, pp. 90–100, Mar. 2006.

[4] J. Wang, E. Chng, C. Xu, H. Lu, and X. Tong, "Identify sports video shots with happy or sad emotions," in *Proc. of ICME*, Toronto, Canada, Jul. 2006.

[5] M. Xu, L.-T. Chia, and J. Jin, "Affective content analysis in comedy and horror videos by audio emotional event detection," in *Proc. of ICME*, Amsterdam, The Netherlands, Jul. 2005.

[6] H.-B. Kang, "Affective content detection using hmms," in *Proc. of ACM MM*, Berkeley, CA, USA, Nov. 2003.

[7] C. T. Castelli, "Trini diagram: imaging emotional identity 3d positioning tool," in *Proc. of SPIE*, Dec. 1999.

[8] B. Brown, *Cinematography, Theory and Practice*, Elsevier, 2002.

[9] J. Itten, *The elements of color*, J. Wiley & Sons, 1970.

[10] J. Van de Weijer, T. Gevers, and A. Gijsenij, "Edge-based color constancy," *IEEE Transactions on Image Processing*, vol. 16, no. 9, pp. 2207–2214, Sep. 2007.

[11] "Cinemetric database," *www.cinemetrics.lv/database*.

[12] "Internet movie database," *www.imdb.com*.